学校代码：10023
学　　号：B2019012012

跨模态图像融合技术在医疗影像分析中的研究

专 业 年 级：北京协和医学院临床医学专业 2019 级试点班

姓　　　名：董袭莹

导　　　师：邱贵兴（教授）
北京协和医学院 临床学院（北京协和医院）
骨科

完 成 日 期：2023 年 5 月

目录

摘要

背景

影像学检查是医疗领域最常用的筛查手段，据统计，医疗数据总量中有超过90%是由影像数据构成[1]。然而，根据亲身参与的临床病例[2]可知，很多情况下，仅凭医生的肉眼观察和主观诊断经验，不足以对影像学异常作一明确判断。而诊断不明引起的频繁就医、贻误病情，则会严重影响患者的生活质量。

相较于传统的主观阅片，人工智能技术通过深度神经网络分析大量影像和诊断数据，学习对病理诊断有用的特征，在客观数据的支持下做出更准确的判断。为了模拟临床医生结合各种成像模式（如 CT、MRI 和 PET）形成诊断的过程，本项目采用跨模态深度学习方法，将各种影像学模态特征进行有机结合，充分利用其各自的独特优势训练深度神经网络，以提高模型性能。鉴于肿瘤相关的影像学资料相对丰富，本项目以宫颈癌和乳腺癌骨转移为例，测试了跨模态深度学习方法在病变区域定位和辅助诊断方面的性能，以解决临床实际问题。

方法

第一部分回顾性纳入了 220 例有 FDG-PET/CT 数据的宫颈癌患者，共计 72,602 张切片图像。应用多种图像预处理策略对 PET 和 CT 图像进行图像增强，并进行感兴趣区域边缘检测、自适应定位和跨模态图像对齐。将对齐后的图像在通道上级联输入目标检测网络进行检测、分析及结果评估。通过与使用单一模态图像及其他 PET-CT 融合方法进行比较，验证本项目提出的 PET-CT 自适应区域特征融合结果在提高模型目标检测性能方面具有显著性优势。第二部分回顾性纳入了 233 例乳腺癌患者，每例样本包含 CT、MRI、或 PET 一至三种模态的全身影像数据，共有 3051 张 CT 切片，3543 张 MRI 切片，1818 张 PET 切片。首先训练 YOLOv5 对每种单一模态图像中的骨转移病灶进行目标检测。根据检测框的置信度划分八个区间，统计每个影像序列不同置信度区间中含有检出骨转移病灶的个数，并以此归一化后作为结构化医疗特征数据，采用级联方式融合三种模态的结构化特征实现跨模态特征融合。再用多种分类模型对结构化数据进行分类和评估。将基于特征转换的跨模态融合数据与特征转换后的单模态结构化数据，以及基于 C3D 分类模型的前融合方式进行比较，验证第二部分提出的方法在乳腺癌骨转移诊断任务中的优越性能。

结果

第一部分的基于跨模态融合的肿瘤检测实验证明，PET-CT 自适应区域特征融合图像显著提高了宫颈癌病变区域检测的准确性。相比使用 CT 或 PET 单模态图像以及其他融合方法生成的多模态图像作为网络输入，目标检测的平均精确度分别

提高了 6.06%和 8.9%，且消除了一些假阳性结果。上述测试结果在使用不同的目标检测模型的情况下保持一致，这表明自适应跨模态融合方法有良好的通用性，可以泛化应用于各种目标检测模型的预处理阶段。第二部分基于特征转换的跨模态病例分类实验证明，跨模态融合数据显著提高了乳腺癌骨转移诊断任务的性能。相较于单模态数据，跨模态融合数据的平均准确率和AUC分别提高了7.9%和8.5%，观察 ROC 曲线和 PR 曲线的形状和面积也具有相同的实验结论：在不同的分类模型中，使用基于特征转换的跨模态数据，相比单模态数据，对于骨转移病例的分类性能更为优越。而相较于基于 C3D 的前融合分类模型，基于特征转换的后融合策略在分类任务方面的性能更优。

结论

本项目主要包含两个部分。第一部分证实了基于区域特征匹配的跨模态图像融合后的数据集在检测性能上优于单模态医学图像数据集和其他融合方法。第二部分提出了一种基于特征转换的跨模态数据融合方法。使用融合后的数据进行分类任务，其分类性能优于仅使用单模态数据进行分类或使用前融合方法的性能。根据不同模态医学图像的特征差异与互补性，本项目验证了跨模态深度学习技术在病变区域定位和辅助诊断方面的优势。相比于只使用单模态数据进行训练的模型，跨模态深度学习技术有更优的诊断准确率，可以有效的成为临床辅助工具，协助和指导临床决策。

关键词：跨模态融合，深度学习，影像分析，宫颈癌，乳腺癌骨转移

Abstract

Background

Imaging examinations serve as the predominant screening method in the medical field. As statistics reveal, imaging data constitute over 90% of the entire medical dataset. Nonetheless, clinical cases have demonstrated that mere subjective diagnoses by clinicians often fall short in making definitive judgments on imaging anomalies. Misdiagnoses or undiagnosed conditions, which result in frequent hospital visits and delayed treatment, can profoundly affect patients' quality of life.

Compared to the traditional subjective image interpretation by clinicians, AI leverages deep neural networks to analyze large-scale imaging and diagnostic data, extracting valuable features for pathology diagnosis, and thus facilitating more accurate decision-making, underpinned by objective data. To emulate clinicians' diagnostic process that integrates various imaging modalities like CT, MRI, and PET, a cross-modal deep learning methodology is employed. This approach synergistically merges features from different imaging modalities, capitalizing on their unique advantages to enhance model performance. Given the ample availability of oncologic imaging data, the project exemplifies the efficacy of this approach in cervical cancer segmentation and detection of breast cancer bone metastasis, thereby addressing pragmatic challenges in clinical practice.

Methods

The first part retrospectively analyzed 72,602 slices of FDG-PET/CT scans from 220 cervical cancer patients. Various preprocessing strategies were applied to enhance PET and CT images, including edge detection, adaptive ROI localization, and cross-modal image fusion. The fused images were then concatenated on a channel-wise basis and fed into the object detection network for the precise segmentation of cervical cancer lesions. Compared to single modality images (either CT or PET) and alternative PET-CT fusion techniques, the proposed method of PET-CT adaptive fusion was found to significantly enhance the object detection performance of the model. The second part of the study retrospectively analyzed 3,051 CT slices, 3,543 MRI slices and 1,818 PET slices from 233 breast cancer patients, with each case containing whole-body imaging of one to three modalities (CT, MRI, or PET). Initially, YOLOv5 was trained to detect bone metastases in images across different modalities. The confidence levels of the prediction boxes were segregated into eight tiers, following which the number of boxes predicting bone metastases in each imaging sequence was tallied within each confidence tier. This count was then normalized

and utilized as a structured feature. The structured features from the three modalities were fused in a cascaded manner for cross-modal fusion. Subsequently, a variety of classification models were then employed to evaluate the structured features for diagnosing bone metastasis. In comparison to feature-transformed single-modal data and the C3D early fusion method, the cross-modal fusion data founded on feature transformation demonstrated superior performance in diagnosing breast cancer bone metastasis.

Results

The first part of our study delivered compelling experimental results, showing a significant improvement in the accuracy of cervical cancer segmentation when using adaptively fused PET-CT images. Our approach outperformed other object detection algorithms based on either single-modal images or multimodal images fused by other methods, with an average accuracy improvement of 6.06% and 8.9%, respectively, while also effectively mitigating false-positive results. These promising test results remained consistent across different object detection models, highlighting the robustness and universality of our adaptive fusion method, which can be generalized in the preprocessing stage of diverse object detection models. The second part of our study demonstrated that cross-modal fusion based on feature transformation could significantly improve the performance of bone metastasis classification models. When compared to algorithms employing single-modal data, models based on cross-modal data had an average increase in accuracy and AUC of 7.9% and 8.5%, respectively. This improvement was further corroborated by the shapes of the ROC and PR curves. Across a range of classification models, feature-transformed cross-modal data consistently outperformed single-modal data in diagnosing breast cancer bone metastasis. Moreover, late fusion strategies grounded in feature transformation exhibited superior performance in classification tasks when juxtaposed with early fusion methods such as C3D.

Conclusions

This project primarily consists of two parts. The first part substantiates that deep learning object detection networks founded on the adaptive cross-modal image fusion method outperform those based on single-modal images or alternative fusion methods. The second part presents a cross-modal fusion approach based on feature transformation. When the fused features are deployed for classification models, they outperform those utilizing solely single-modal data or the early fusion model. In light of the differences and complementarity in the features of various image modalities, this project underscores the strengths of cross-modal deep learning in lesion segmentation and disease classification.